

## LECTURA Y ESCRITURA CON HERRAMIENTAS BASADAS EN ANOTACIÓN DE TEXTOS DIGITALIZADOS **Entrevista a los grupos de investigación L.E.E.T.HI e I.L.S.A.**

### Datos de la entrevista

*Entrevistadoras: M<sup>a</sup> Carmen Gertrudis Casado y M<sup>a</sup> Carmen Gálvez de la Cuesta (Investigadoras del Grupo Ciberimaginario URJC-Icono14)*

*Entrevistados: Amelia Sanz Cabrerizo y José Luis Sierra Rodríguez (Grupos de Investigación de L.E.E.T.HI e I.L.S.A.)*

### **M<sup>a</sup> Carmen Gertrudis Casado**

Investigadora

Grupo Ciberimaginario (URJC-Icono14). Facultad de Ciencias de la Comunicación. Universidad Rey Juan Carlos. Camino del Molino, s/n Fuenlabrada, 28963 Madrid (España) - Email: [carmen.gertrudis@ciberimaginario.es](mailto:carmen.gertrudis@ciberimaginario.es)

### **M<sup>a</sup> Carmen Gálvez de la Cuesta**

Investigadora

Grupo Ciberimaginario (URJC-Icono14) Facultad de Ciencias de la Comunicación. Universidad Rey Juan Carlos. Camino del Molino, s/n Fuenlabrada, 28963 Madrid (España) - Email: [carmen.galvez@ciberimaginario.es](mailto:carmen.galvez@ciberimaginario.es)

## Presentación de los entrevistados

**Amelia Sanz Cabrerizo** es Profesora Titular en el Departamento de Filología Francesa de la UCM. Doctora en Filología por la Universidad Complutense (1991), Licenciada en Filología Española por la Universidad Complutense y Licenciada en Filología Francesa por la Universidad Nacy (Francia). De 1987 a 1991 fue Profesora Adjunta de Español en la Universidad Jean Monnet (Saint-Etienne, Francia). En 1991 se incorporó al Departamento de Filología Francesa, de la UCM, donde es Profesora Titular de Literatura francesa desde 2007. Desde 2000 dirige el grupo de investigación LEETHI (Españolas y Europeas literaturas: Del Texto al hipermedia - Literaturas españolas y europeas: del texto al hipermedia). En la actualidad, La Profesora Sanz es la Directora del Campus Virtual de la UCM.

En el grupo LEETHI, la Profesora Sanz ha liderado varios proyectos sobre aspectos relacionados con el análisis de textos, lectura de textos electrónicos, la digitalización y anotación de textos literarios. Los resultados de su investigación aparecen en varios trabajos de investigación y monografías. Ha participado como revisora en varias revistas internacionales (Theleme. Revista Complutense de Estudios Franceses), editoriales (Arco Libros) y

conferencias internacionales (XIX Internacional de Literatura Comparada Congreso de la Asociación, Seúl 2010). Participa a su vez, en el proyecto europeo COST: Los Escritores de la Mujer en la Historia, como miembro del Comité de Gestión, con otros investigadores procedentes de 18 países europeos. Sus intereses de investigación incluyen la literatura comparada, la interculturalidad, intermedialidad, las humanidades digitales y la literatura digital.

Por su parte, **José Luis Sierra Rodríguez** es Profesor Titular de Universidad en el Departamento de Ingeniería del Software e Inteligencia Artificial de la Universidad Complutense de Madrid (España). Es Diplomado (Escuela Universitaria de Informática de la Universidad Politécnica de Madrid, 1991) y Licenciado en Informática (Facultad de Informática de la Universidad Politécnica de Madrid, 1995). Tiene un Máster en Tecnologías de la Información en Fabricación (Universidad Politécnica de Madrid, 1992), y es Doctor en Informática por la Universidad Complutense de Madrid (2004). Hasta 1998 fue asistente de investigación en el Instituto Industrial de Automática del Consejo de Investigaciones Científicas (1990-1992) y en el Departamento de Inteligencia Artifi-

cial de la UPM (1993-1998). En 1998 se incorporó a la UCM como profesor auxiliar (Profesor Ayudante), donde es Profesor Titular en Ciencias de la Computación desde 2007.

En la UCM, el Dr. Sierra lidera el Grupo de Investigación de ILSA (Grupo de Investigación en Ingeniería de lenguajes de software Aplicaciones y). Ha liderado y participado en varios proyectos de investigación en los campos de las humanidades digitales, e-learning de software y procesamiento del lenguaje, cuyos resultados han sido

publicados en más de 100 trabajos de investigación en revistas internacionales, conferencias y capítulos de libros. Colabora regularmente como miembro revisor en varias revistas internacionales (por ejemplo, IEEE Software, Ciencias de la Información, interactuar con los ordenadores) y conferencias (por ejemplo, ICALT, e-Learning, WAPL). Sus intereses de investigación se centran en el desarrollo y los usos prácticos del lenguaje informático.

**Gráfico n° 1: Amelia Sanz Cabrerizo y José Luis Sierra Rodríguez**



## Entrevista

*¿Podrías exponer brevemente los objetivos y líneas de investigación del Grupo de Investigación L.E.E.T.HI (Literaturas Europeas del Texto al Hipermedia)?*

**Amelia. LEETHI** lleva 10 años trabajando sobre los cambios graduales que se están produciendo en la lectura/escritura literaria de la página en papel a la pantalla electrónica en nuestro siglo. El Grupo tiene unas características específicas: su carácter intercultural (con miembros que provienen de diferentes lenguas y literaturas), intergeneracional (desde “inmigrantes” a “nativos digitales”) e interdisciplinar (investigadores de Filología y Educación y ahora por la colaboración con el Grupo ILSA de la Facultad de Informática).

Comenzamos a trabajar sobre este tema porque nos dimos cuenta de que no podíamos enseñar las literaturas de la misma manera que hicieron nuestros maestros porque la definición misma de lo literario ha cambiado (tanto el estatuto social de las literaturas, como su soporte), sino por el perfil de nuestros lectores (y escritores): la llamada “Generación Google”<sup>1</sup>. Desde luego, cuando las bibliotecas digitalizan sus fondos literarios, los profesores de literatura deben determinar qué textos

dar a leer a sus estudiantes, cómo y para qué.

Creemos que los medios electrónicos abren posibilidades: es el momento de aprovechar la nueva “patrimonialización” de las literaturas que están realizando las instituciones mediante la digitalización masiva, porque, pensando que las literaturas siguen siendo consideradas el corpus creador de identidad regional/ nacional/ continental, ahora podemos acceder y dar a leer textos escondidos y olvidados, esto es, romper el canon; es el momento de aprovechar las nuevas condiciones de la cientificidad, esto es, servirnos de las utilidades de las ciencias “beta” que son reproductivas para dar respuesta a las preguntas tradicionales de las ciencias “alfa” que son interpretativas, en busca de nuevo conocimiento; esto es, es tiempo de enseñar a leer las literaturas con TICC (Tecnologías de la Información y de la Cultura); es el momento de aprovechar las otras modalidades de lectura posibles que nos brindan estas tecnologías, aunque se anuncie la muerte de la lectura reflexiva frente al triunfo de una lectura superficial, porque las pantallas (planas, horizontales, ligeras y móviles) o la tinta electrónica están favoreciendo otros rituales de lectura y hemos de explorarlos. Son estos plan-

teamientos los que nos han llevado a colaborar con el equipo de José Luis Sierra: ILSA.

***¿Podrías exponer brevemente los objetivos y líneas de investigación del Grupo de Investigación I.L.S.A. (Ingeniería de Lenguajes Software y Aplicaciones)?***

**José Luis.** El grupo de Investigación ILSA (Ingeniería de Lenguajes Software y Aplicaciones; [ilsa.fdi.ucm.es](http://ilsa.fdi.ucm.es)) es un grupo de relativamente reciente creación, que nace en la Facultad de Informática de la Universidad Complutense de Madrid en el año 2010, y que, en 2011, es validado positivamente y reconocido oficialmente por la Universidad Complutense de Madrid (grupo no. 962022). La línea de investigación básica que da origen al grupo se encuentra en los trabajos de investigación sobre Desarrollo de Software Educativo Dirigido por Lenguajes iniciados por mi mentor, el Profesor Alfredo Fernández-Valmayor Crespo, quién, desafortunadamente para nosotros, se jubiló en 2010. No obstante, Alfredo sigue haciendo bastantes cosas (por ejemplo, ahora está promoviendo la consolidación de una Fundación para la preservación del Patrimonio Arqueológico de Panamá: <http://www.fundacionelcano.org/>), y, desde la distancia, continúa manteniendo con nosotros una relación fruc-

tífera, tanto en el plano científico como en el personal.

La idea del desarrollo dirigido por lenguajes es simple: proporcionar a los expertos en el dominio (filólogos, arqueólogos, historiadores, profesores, etc.), en lugar de aplicaciones cerradas, lenguajes informáticos de propósito específico que les permitan configurar y adaptar las aplicaciones a sus necesidades. De esta forma, en ILSA trabajamos en dos frentes diferentes. Por una parte, aplicamos los principios del desarrollo de software dirigido por lenguajes a los ámbitos del aprendizaje electrónico (e-learning) y de las humanidades digitales, lo que nos ha llevado a establecer una fructífera colaboración con el grupo LEETHI. Por otra parte, llevamos a cabo una investigación en aspectos Informáticos más fundamentales, orientada a idear mecanismos que nos permitan facilitar y sistematizar dicho desarrollo dirigido por lenguajes. En este sentido, estamos aplicando técnicas utilizadas por los Informáticos en la creación de entornos de programación a distintos aspectos del desarrollo de aplicaciones en el campo del e-learning y de las humanidades digitales: procesamiento eficiente de información semiestructurada (tal como, por ejemplo, documentos XML, o bancos de datos vinculados RDF), transformación de las estructuras de información internas manejadas por

los programas informáticos (transformación de modelos), y especificación formal de la interacción en aplicaciones interactivas y aplicaciones web. En concreto, los instrumentos concretos en las que nos centramos para abordar todos estos aspectos del desarrollo se denominan *gramáticas de atributos*. Son modelos formales del lenguaje que tienen ya bastante solera en informática (fueron propuestos, allá por los finales de los años sesenta del siglo pasado, por el Prof. Donald D. Knuth, uno de los pioneros de la Informática, inventor, entre otras cosas, del popular sistema de edición de textos TeX, que de seguro sonara a todo linuxero que se precie), y a los que, en ILSA, esperamos dar nuevas utilidades en el desarrollo sistemático de Software.

***Habladnos del proyecto: ¿en qué consiste el proyecto Lectura y escritura con herramientas basadas en anotación de textos digitalizados?***

**Amelia.** Como ya sabes, la Universidad Complutense fue una de las primeras universidades europeas en colaborar con el proyecto Google Books ([Biblioteca Complutense - Google](#)) cuando se propuso poner en la Red los 100.000 volúmenes que constituyen su colección del fondo antiguo en la Biblioteca Marqués de Valdecilla gracias a las tecnologías de Google. Pero a los profesores de literatura nos parece que todos estos textos digitali-

zados no serán de verdadera utilidad en literatura si no resultan atractivos, asequibles, accesibles a los estudiantes lectores. Y lo cierto es que muchos de los textos digitalizados no parecen fácilmente legibles y hasta resultan inexplotables para lo que nosotros consideramos la lectura académica, una lectura de segundo grado.

Los profesores, para enseñar, necesitamos textos de fácil lectura en cuatro niveles: una documentación completa sobre el volumen, una perfecta reproducción en modo imagen y en modo texto, una presentación de textos enriquecidos gracias a notas insertadas, un uso educativo de los textos de forma que se facilite el acceso a lectores poco experimentados en la lectura y, más aún, en la lectura electrónica. A menos que se creen estas herramientas, corremos el riesgo de que proyectos de digitalización masiva como Google Books sean ignorados para nuestros estudiantes y, con ello, el patrimonio literario.

En este camino, LEETHI e ILSA nos atrevimos a lanzar el proyecto *Collaborative Annotation of Digitalized Literary Texts* que presentamos a comienzos de 2010 a los *Google Awards in Digital Humanities*, un proyecto cuya financiación nos fue concedida y renovada en 2011 para crear un método de anotaciones innovador, especialmente concebido para la colección

de textos digitalizados por el UCM-Google, en un entorno amigable que facilite esa lectura de segundo grado que perseguimos en la Universidad.

Este sistema de anotación permite anotar textos y clasificar esas notas de forma colaborativa mediante esquemas de anotaciones. Para ello, el sistema incluye las siguientes funcionalidades:

- recuperación de textos digitalizados en la Red y anotación de éstos;
- establecimiento de términos y relaciones que conformen el esquema de anotaciones según son creadas esas anotaciones;
- búsqueda y navegación dentro de las anotaciones utilizando las estructuras del esquema;
- edición y reestructuración del esquema por parte del usuario;
- realización de tareas administrativas, como creación de grupos de trabajo, concesión de privilegios, etc.

Creemos que nuestro sistema de anotaciones posee una serie de puntos fuertes, como son:

- la creación colaborativa de esquemas de anotaciones, de forma que evitemos los inconvenientes que presentan los esquemas de anotaciones pre-establecidas, con frecuencia demasiado complejos y, sin embargo, siempre insuficientes;

- la creación de un metamodelo de anotaciones altamente estructurado a partir de modelos lingüísticos como son los tesauros y las ontologías formales, con el fin de colmar la falta de estructura que suelen presentar los métodos de etiquetado basados en palabras clave;

- la incorporación de mecanismos de revisión que permitan al usuario reestructurar los términos y las relaciones a fin de resolver las incongruencias y mejorar la calidad del esquema de anotación, lo cual nos permite resolver los problemas lingüísticos que nos plantean plurales, polisemias, sinonimias y homonimias.

**José Luis.** La verdad es que estamos bastante satisfechos con los resultados del proyecto. La financiación brindada por Google nos ha permitido crear un sistema francamente interesante, utilizando tecnologías punteras en el desarrollo de aplicaciones web altamente interactivas (aplicaciones RIA, de *Rich-Internet Applications*). Me estoy refiriendo a las tecnologías de Google GWT (Google-Web Toolkit), y AE (Application Engine), la infraestructura de Google para la computación en la nube). Así mismo, en los próximos meses vamos a llevar a cabo algunos experimentos francamente interesantes desde el punto de vista informático, orientados a introducir las gramáticas



de atributos en el proceso de desarrollo. Básicamente lo que queremos hacer es componentizar la aplicación en términos de servicios, y utilizar gramáticas de atributos para orquestar dichos servicios.

Me gustaría, así mismo, resaltar aquí *lo bien* que hemos trabajado, y que continuamos trabajando, en este proyecto, debido básicamente al modelo *americano* de financiación de proyectos de investigación seguido en Google. Para que te hagas una idea, Google dice: “aquí tenéis el dinero... investigad como queráis; simplemente, cuando terminéis enseñarnos lo que habéis hecho; si nos gusta, os damos más, y si no... tan amigos”. Nada de deadlines, nada de papeleos, informes inútiles, producción de montañas y montañas de papel... simplemente hacer lo que sabemos hacer: investigar, y construir cosas interesantes. Esto contrasta, por ejemplo, con los proyectos competitivos nacionales o de la Comunidad Europea, que, como ya sabrás, son un infierno de burocracia, papeleo y demás. La verdad es que los sistemas de financiación de la investigación nacionales y europeos tienen mucho que aprender todavía...

**¿A quién van dirigidas estas herramientas?**

**Amelia.** Desde el comienzo concebimos la herramienta en dos dimensiones: habría de servir para enseñar a

leer y convertir entonces el texto en objeto de aprendizaje; habría de dar servicio a los especialistas para la realización de ediciones críticas anotadas, lo cual es de interés para las editoriales. No queremos olvidar en ningún momento que pretendemos es enseñar a leer, esto es, enseñar a pensar.

Creemos que las aportaciones de un sistema de anotación como éste son muchas para los profesores de literatura: ponemos a disposición de profesores y estudiantes un gran número de textos literarios fuera del canon nacional que pueden ser leídos, anotados e incorporados así a una lectura académica más plural; exploramos un modo de lectura y anotación colaborativa que convierta esta experiencia en global, en la Red, y que, al tiempo, permite crear comunidades de usuarios comprometidos en el desarrollo de los patrimonios literarios; estudiamos los rituales de lectura y de escritura colaborativa que los usuarios tienen que desarrollar ante la pantalla, de forma que podamos aprovechar positivamente la experiencia para saltar de un modo de lectura-escritura en papel a otras modalidades electrónicas; insertamos estos textos anotados con @Note y los trabajos realizados con ellos a la dinámica del aprendizaje mediante su incorporación a cualquier entorno de aprendizaje virtual, como Moodle o Sakai.



Queremos que sea una contribución más a lo que ya es, de hecho, el renacimiento del lecto-escritor.

**José Luis.** La verdad que la confluencia de comunidades con experiencias tan dispares como las que se han encontrado en el desarrollo de @note es muy estimulante desde el punto de vista informático. Aquí lo difícil no es utilizar la última tecnología (eso los informáticos los sabemos hacer muy bien). Aquí lo difícil es entender qué es lo que realmente necesitan los expertos en el dominio (en este caso, aquí nuestros colegas filólogos). Para que el desarrollo de este tipo de proyecto tenga éxito es necesario involucrar activamente a los expertos, que son realmente los que saben, los que conocen el dominio, y es aquí donde, en mi opinión, los métodos de desarrollo dirigido por lenguajes pueden ser una herramienta valiosísima.

**¿Crees que la Web 2.0 ha generado nuevas formas de creación literaria?**

**Amelia.** Sin duda, pero creo que tu pregunta se refiere a la literatura digital, esto es, aquella literatura creada por y para los sistemas electrónicos, esto es, que no se puede trasvasar al papel sin que se pierda. Es el objeto del Programa Erasmus Intensivo que estamos desarrollando en la Universidad Complutense: *European Digital Literatures*

(<http://www.digitalliterature.eu>). Nuestra herramienta sirve para anotar la literatura digitalizada, esto es, aquella que simplemente ha pasado del soporte papel, pero pronto también servirá para anotar esta nueva literatura.

En cualquier caso, y como ya he dicho en muchas ocasiones anteriores, estoy segura de que la literatura será digital o no será.

**Cita de este artículo**

Gálvez de la Cuesta, M<sup>a</sup> C. y Gertrudis Casado, M<sup>a</sup> C. (2012) Lectura y escritura con herramientas basadas en anotación de textos digitalizados. Entrevista a los grupos de investigación L.E.E.T.HI e I.L.S.A. *Icono14* 10(2), 6-28, doi: [10.7195/ri14.v10i2.503](https://doi.org/10.7195/ri14.v10i2.503)

---

## NOTAS

<sup>i</sup> Para una discusión sobre la “Google Generation”, véase CIBER, *Information Behaviour of the Researcher of the future*, in <http://www.bl.uk/news/pdf/googlegen.pdf>. Consultado en septiembre 2009.